AIDS in Africa

Editors

Max Essex, DVM, PhD
Chairman
Harvard AIDS Institute
Mary Woodard Lasker Professor of Health Sciences
Harvard University
Boston, Massachusetts, USA

Souleymane Mboup, PharmD, PhD
Professor of Bacteriology and Virology
Laboratory of Bacteriology and Virology
School of Medicine and Pharmacy
Université Cheikh Anta Diop
Dakar, Senegal

Phyllis J. Kanki, DVM, SD
Associate Professor of Pathobiology
Department of Cancer Biology
Harvard School of Public Health
Boston, Massachusetts, USA

Mbowa R. Kalengayi, MD, DSc
Ordinarius Professor and Chairman
Department of Pathology
University of Kinshasa Medical School
Kinshasa, Zaire

Managing Editor

Paula J. Brewer, MA
Manager of Communications
Harvard AIDS Institute
Cambridge, Massachusetts, USA

Raven Press **New York**

Raven Press, Ltd., 1185 Avenue of the Americas, New York, New York 10036

Made in the United States of America

Library of Congress Cataloging-in-Publication Data

AIDS in Africa / managing editor, Paula J. Brewer ; editors, Max Essex . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-7817-0110-4
1. AIDS (Disease)—Africa. I. Essex, Myron.
[DNLM: 1. Acquired Immunodeficiency Syndrome—epidemiology—Africa. 2. Acquired Immunodeficiency Syndrome—therapy. 3. HIV Seroprevalence. 4. Socioeconomic Factors. WD 308 A28831534 1993]
RA644.A25A3634 1993
362.1'969792'0096—dc20
DNLM/DLC
for Library of Congress 93-24662

The material contained in this volume was submitted as previously unpublished material, except in the instance in which some of the illustrative material was derived.

Great care has been taken to maintain the accuracy of the information contained in the volume. However, neither Raven Press nor the editors can be held responsible for errors or for any consequence arising from the use of the information contained herein.

Material appearing in this book prepared by individuals as part of their official duties as U.S. Government employees are not covered by the above-mentioned copyright.

9 8 7 6 5 4 3 2 1

AIDS in Africa, edited by Max Essex, Souleymane Mboup, Phyllis J. Kanki, and Mbowa R. Kalengayi.
Raven Press, Ltd., New York © 1994.

2

The Molecular Biology of HIV-1

Jean-Pierre Vartanian and Simon Wain-Hobson

Unitè de Rétrovirologie Moléculaire, Institut Pasteur, 75724 Paris cedex 15, France.

No group of RNA viruses has as diverse a genetic organization or as plastic a genome as the retroviruses. Their ability to transduce, or steal, genetic material from the host cell genome—a remarkable feat of genetic engineering—gave rise to the oncogene, a discovery that has revolutionized cancer research. With the identification of HIV as the etiologic agent of AIDS, retrovirology entered its renaissance. Many retroviruses may still be discovered, just as HIV emerged unexpectedly and dramatically. The recent description of a novel retrovirus associated with solid tumors in fish is just such an example (1).

Retroviruses have been conventionally divided into three subfamilies based on biologic and morphologic criteria: oncoviruses, spumaviruses, and lentiviruses. This classification has since become outdated, as retroviruses have been found to be more complex and varied than first thought.

The oncoviruses are tumor-promoting viruses that can be divided into three or four distinct groups, not all of which induce malignancies. For example, the oncogenic potential of the Mason-Pfizer monkey virus (MPMV), a D-type retrovirus, was sought without success (2). Yet the immunosuppressive effect of the virus in juvenile monkeys was overlooked. It was not until 1984, when an MPMV-like D-type retrovirus was isolated from rhesus monkeys with profound immunosuppression, that this aspect of D-type retroviruses became apparent (3).

Until recently, little was known about the spumaviruses, or foamy viruses. Although more is being learned about these viruses, an association with disease has yet to be established (4). Clearly, retroviruses are not always pathogenic.

With the extensive sequence data now available, it has been possible to construct a phylogenetic tree for the retrovirus family (Fig. 1). Certain groups—such as the human T-cell lymphotropic virus (HTLV) group of oncoviruses and the spumaviruses—are well defined. Other groups, such as the lentiviruses, are diverse. Even among the primate lentiviruses, the genetic organization of other parts of the genome is highly variable (Fig. 2). Recombination between retroviruses of two different groups can occur, further complicating classification (5).

Finally, it is important not to forget the endogenous retroviruses. These are proviruses, usually but not invariably defective, that replicate along with the host cell genome. There are examples akin to the murine and avian leukemia viruses as well as to the B-type and D-type retroviruses. Endogenous retroviruses homologous to any of the lentiviruses, spumaviruses, or HTLV group do not appear to exist.

Lentiviruses, named for their long incubation period between infection and overt

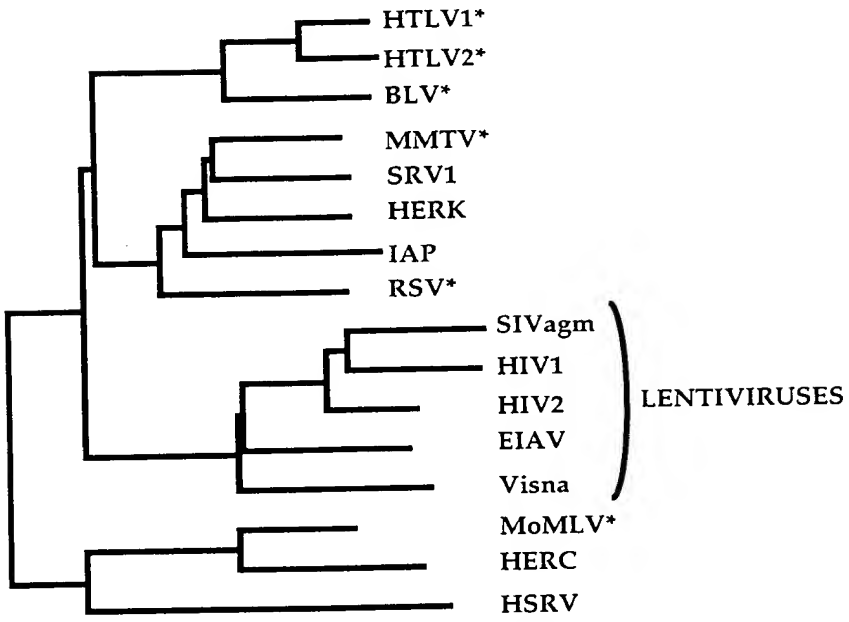

FIG. 1. A phylogeny of retroviruses based on reverse transcriptase sequences. An asterisk denotes tumor association. HTLV-1 and HTLV-2, human T-cell lymphotropic virus types 1 and 2; BLV, bovine leukemia virus; MMTV, mouse mammary tumor virus; HERK, human endogenous retrovirus K family; IAP, hamster intracisternal A-type particle; RSV, Rous sarcoma virus; SIV_{agm}, simian immunodeficiency virus from African green monkeys; HIV-1 and HIV-2, human immunodeficiency virus types 1 and 2; EIAV, equine infectious anemia virus; Visna, ovine maedi/visna virus; MoMLV, Moloney murine leukemia virus; HERC, human endogenous retrovirus C-type family; HSRV, human spuma or foamy retrovirus. Adapted from ref. 125.

disease, have, until AIDS, been neglected despite their association with neurologic, pulmonary, arthritic, and hematologic diseases in outbred animal populations (6). The prototype lentivirus—the visna virus—was first reported in 1960 as the cause of an epidemic of deaths among Icelandic sheep in the 1930s and 1940s (7). Visna, which means "wasting" in Icelandic, primarily causes neurologic disease and pulmonary lesions. Lentiviruses were later discovered in goats—the caprine arthritis encephalitis virus—and in horses—the equine infectious anemia virus (8,9). The former provokes leukoencephalopathy in kid goats and chronic arthritis in adult goats; the latter causes anemia in horses. Their lack of association with tumors led lentiviruses to be neglected, however. For example, the bovine immunodeficiency virus, a milk-borne lentivirus, was first isolated in 1972, yet ignored until 1987, after the AIDS viruses had been well established (10).

Lentivirology changed dramatically with the discovery of HIV-1 (11) in 1983 (Table 1). Within five years the simian immunodeficiency virus (SIV) had been isolated from macaques (12), African green monkeys (13), mandrills (14), sooty mangabeys (15,16), and chimpanzees (17). A second human virus, HIV-2 (18,19), was identified in West Africa, and the discovery of the feline immunodeficiency virus, a widely diffused virus of cats, soon followed (20). It is likely that more lentiviruses will be

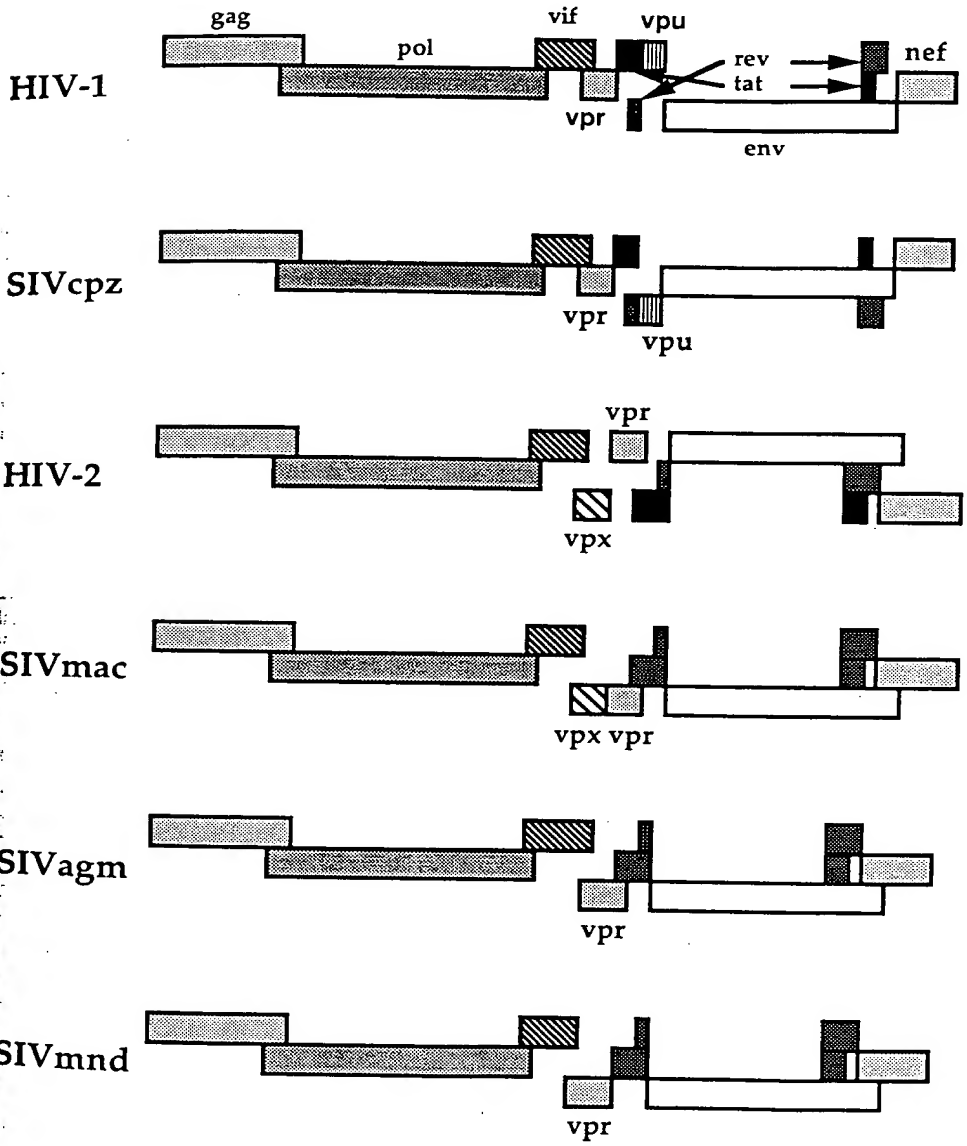

FIG. 2. Organization of the coding potential of primate immunodeficiency viruses. The LTR structures that flank the orfs have been omitted for clarity. The *vpu* orf specific to HIV-1 and SIV_{cpz} and the *vpx* orf specific to HIV-2 and SIV_{mac} are worth noting. The abbreviations are the same as those in Tables 1 and 3.

discovered in the years to come. Ironically, lentivirology has begun to eclipse research on other retroviruses, and AIDS-related research is now a driving force in biomedical research. (See Table 2 for a general comparison of characteristics of several human retroviruses.)

The few lines devoted here to these viruses are not simply for historic interest, for

TABLE 1. *Mammalian lentiviruses*

Lentivirus	Disease	Reference
Ovine maedi/visna virus (OMVV) or visna virus	Pulmonary (maedi) and neurologic (visna)	7
Caprine arthritis encephalitis virus (CAEV)	Leukoencephalopathy in kid goats and arthritis in adult goats	8
Equine infectious anemia virus (EIAV)	Infectious anemia	9
Bovine immunodeficiency virus (BIV)	Unknown	10
Human immunodeficiency virus type 1 (HIV-1)	AIDS	11
Simian immunodeficiency virus of captive macaques (SIV_{mac})	AIDS	12
Human immunodeficiency virus type 2 (HIV-2)	AIDS	18,19
Simian immunodeficiency virus of sooty mangabeys (SIV_{sm})	Unapparent disease so far[a]	15,16
Feline immunodeficiency virus (FIV)	AIDS	20
Simian immunodeficiency virus of African green monkeys (SIV_{agm})	Unapparent disease so far	13
Simian immunodeficiency virus of mandrills (SIV_{mnd})	Unapparent disease so far	14
Simian immunodeficiency virus of chimpanzees (SIV_{cpz})	Unapparent disease so far	17

[a]"Unapparent disease so far" does not mean that the viruses are apathogenic, but that data gathered either in the wild or with adequate follow-up time in captivity are insufficient to discern the pathogenic potential of these viruses.

in comparing their genetic structures (Fig. 2) we may start to appreciate the extraordinary plasticity of the retroviral genome. The comparison also helps us to realize which viruses constitute genetically correct models of human lentiviruses. Thus, none of the ungulate viruses has as complex a genetic structure as HIV-1 and HIV-2, demonstrating that, although they may provide powerful insights into the mech-

TABLE 2. *Characteristics of various human retroviruses*

Retrovirus	Subfamily	Clinical Characteristics	Major Target Cell	Characteristics of Virus Particle
HTLV-I	Oncovirinae	ATL HAM/TSP	CD4(+) T cell (non-lymphoid cells *in vitro*)	80–120 nm enveloped spherical electron-dense core
HTLV-II	Oncovirinae	Hairy cell leukemia (?)	CD8(+) T cell *in vivo*	80–120 nm enveloped spherical electron-dense core
HFV	Spumavirinae	None established	Multiple cell types	100–140 nm enveloped spherical core
HIV-1	Lentivirinae	Immunodeficiency	CD4(+) T cell monocyte/ macrophage	80–120 nm enveloped cylindrical electron-dense core
HIV-2	Lentivirinae	Immunodeficiency	CD4(+) T cell monocyte/ macrophage	80–120 nm enveloped cylindrical electron-dense core

ATL = adult T-cell leukemia; HAM/TSP = HTLV-I-associated myelopathy/tropical spastic paraparesis; HFV = human foamy virus

anisms of lentiviral disease, their value is somewhat circumscribed. Indeed, the incubation time from infection to overt clinical disease may take an animal's lifetime. The signal importance of the macrophage in these infections was first documented for the ungulate lentiviruses (21).

The chimpanzee lentivirus, SIV_{cpz} (22), is the primate counterpart of HIV-1, whereas lentiviruses in the macaque (SIV_{mac}) and sooty mangabey (SIV_{sm}) are closest to HIV-2 (23,24). Given the problems inherent in using chimpanzees for experimental research, including their paucity, we are effectively without a manageable animal model for HIV-1 infection. The $SIV_{mac/sm}$ model and the HIV-2–adapted infection of macaques clearly hold tremendous promise for ultimately understanding the HIVs and AIDS pathogenesis. (See Chapter 5, *this volume*.)

We will devote the rest of the chapter to the biology of HIV-1. (Throughout the chapter "HIV" will refer to HIV-1 unless specified as HIV-2.) We will discuss virus-host interactions, the replicative strategy of the virus, the assembly of the virion, and the tremendous genetic diversity of the virus.

THE VIRUS AND THE CELL

Cell Tropism

Retroviruses are enveloped viruses with two copies of encapsidated plus stranded genomic RNA. The lipid membrane is of cellular origin and is acquired during the budding of the virus from the cell surface. Table 3 shows the viral proteins along with their principal functions. Figure 2 illustrates the position of the gene encoding each protein and Fig. 3B illustrates the position of each protein in virus particles.

HIV has two virions: the "conventional" virion (Fig. 3A and B) and the latently infected lymphocyte. Both are probably transmitted in all forms of infection, with the exception of cell-free blood products. The proportion of cell-free virus in the blood is substantial just before seroconversion, after which it declines to low levels (25,26). The titers then rise with declining clinical status. Interestingly, much cell-free virus is associated with antibody. The relative contribution of cell-free or cell-associated virus to natural infection is not known. The latently infected lymphocyte is activated in the new host by a mixed lymphocyte reaction resulting in the production of numerous conventional virions.

A combination of studies based on polymerase chain reaction (PCR) and *in situ* hybridization has shown that whereas the number of peripheral blood mononuclear cells (PBMCs) supporting active replication is of the order of 1/100,000 (27), the number of latently infected PBMCs can vary from 1/50,000 to 1/100 depending on disease stage (28,29). *In vivo*, HIV is found essentially in $CD4^+$ lymphocytes, and then mainly within $CD4^+$ memory cells (30). Antigen-presenting cells, monocytes, macrophages, Langerhans cells, Kupffer cells, and microglial cells also can be infected. Although there are some reports of other cell types naturally harboring HIV, there is no good consensus on this point. *In vitro*, however, HIV can infect cell lines of lymphoid, hepatocyte, and fibroblast types, suggesting another mechanism for virus-cell interaction and infection. A novel galactocerebroside molecule has been identified as the receptor for HIV on $CD4^-$ glial and neuroblastoma cells (31). The dissociation constant (Kd) for this interaction was surprisingly low (10^{-9}M). The relevance of the preceding *in vitro* data to natural infection is not clear, however, and these data probably should be treated with caution. The viral long terminal re-

TABLE 3. *HIV proteins and their known characteristics and functions*

Structural	
Gag Pr55	Structural nucleocapsid precursor; NH_2 myristylated, directing it to inner cytoplasmic membrane; cleaved by viral protease
p17MA	Gag matrix protein; core protein anchoring virion envelope via NH_2 myristylated terminus.
p24CA	Gag capsid protein within the p17 shell
p14NC	Gag nucleic acid binding protein; presumably condenses RNA genomes within virion; frequently cleaved into p7 and p6, with p7 retaining binding capacity
Env gp160	Highly glycosylated precursor to env products
gp120	Surface glycoprotein, >50% carbohydrate; binds to HIV receptor $CD4^+$ molecule
gp41	Transmembrane protein; amino terminus fuses with plasma membrane
Enzymatic	
Gag-Pol Pr180	Precursor encoding viral enzymes in form of polyprotein; cleaved *pol* in budding and immature complete virions
p12PR	Viral proteinase belonging to the aspartic acid group; ensures cleavage of the gag and gag-pol precursors; functions as a dimer
p66RT	Mature reverse transcriptase/RNaseH; functions as a dimer; one subunit is cleaved by the viral protease, leaving a functional p66:51 heterodimer
p32IN	Viral endonuclease/integrase results in integration of the provirus
Regulatory	
Tat p16/p14	Viral transactivator; nuclear/nucleolar localization; binds TAR
Rev p19	RNA transport phosphoprotein, primarily nucleolar; binds RRE
Accessory	
Vif p23	Viral infectivity factor
Vpr p10-15	Virion associated
Vpu p16	Unique to HIV-1; helps intracellular transport of gp160 through Golgi and promotes release of virions of regular morphology
Nef p27	Myristylated phosphoprotein associated with inner cytoplasmic membrane; necessary for high viral load *in vivo*
Vpx p12-16	Virion-associated protein unique to HIV-2

TAR, *tat* response element; RRE, *rev*-responsive element.

peat (LTR) does not appear to play a role in determining cell tropism as was postulated for a variety of murine retroviruses.

Receptor

The receptor for HIV—and for all the primate immunodeficiency viruses—is the $CD4^+$ molecule (32,33), a 55-kDa surface glycoprotein belonging to the immunoglobulin superfamily (34). $CD4^+$ is recognized by the HIV surface glycoprotein gp120, an interaction that is extremely specific and tight yet depends on the isolate used. Thus, the Kd is of the order of 4×10^{-9}M (35). The three-dimensional structure of $CD4^+$ is known (36,37), and the major residues involved in binding have been identified. They essentially map to residues 37 to 53, which form a loop at the surface of the membrane distal V1 domain (38,39).

Several hypotheses have been proposed for the entry of HIV into cells: the fusion of viral and cellular membrane in a pH-independent manner (40) by envelope fusion with the cell membrane; receptor-mediated endocytosis (41); or both fusion and endocytosis. At some point fusion of the virion and host cell membranes or vesicles occurs, mediated by the hydrophobic terminus of the viral transmembrane protein gp41. Thus, recognition of $CD4^+$ is essential yet insufficient for infection.

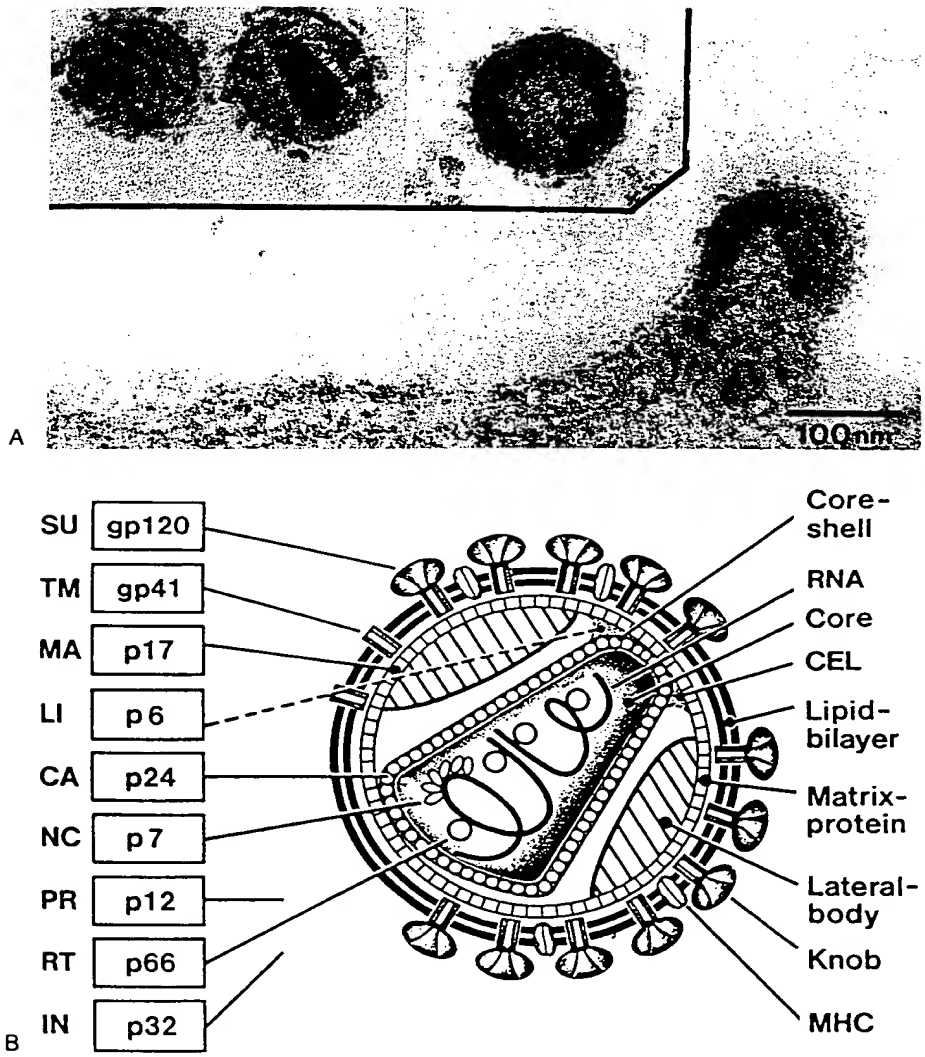

FIG. 3. The HIV-1 virion. **A:** Electron micrographs of HIV-1 budding from the cell surface membrane. The inset shows a perfectly symmetrical immature virion and a pair of mature, and hence infectious, virions. **B:** Sketch of the mature HIV-1 virion. The origin of the lateral bodies remains obscure. The core envelope link (CEL) may be mediated by p6 (LI link protein), although this is not proved. From ref. 103.

Human T lymphocytes exist as long-term resting small lymphocytes and short-lived activated blasts. HIV infects only the latter, which are activated by mitogens *in vitro* and presumably by antigens *in vivo*. Of the few experiments reporting infection of resting lymphocytes, extensive use was made of ultrasensitive PCR methods (42). It is difficult not to believe that a few donor PBMCs were dividing when blood was drawn or that preparation of the PBMCs did not activate a few lymphocytes.

The situation is different when it comes to antigen-presenting cells. HIV can enter

cells via the Fc receptor by internalization of antibody/virus complexes. Antibody-enhanced infection could be blocked by monoclonal antibodies to FcRIII (43,44). HIV probably also enters via phagocytosis. The important point to retain is that antigen-presenting cells can produce HIV constitutively without the need for external stimuli probably because of activation of the precursor form of the NF-κB transcription enhancer by the viral protease (45).

THE PROVIRUS

Formation of the Provirus

Once within the cytoplasm, reverse transcription of the single-stranded virion RNA into double-stranded proviral DNA occurs within a permeable core-like structure. The synthesis of the minus strand of DNA is initiated near the 5′ end of the plus strand RNA by a specific tRNA primer, tRNAlys3 in the case of HIV-1, co-packaged in the viral particles, and it binds via its 3′ end to a complementary sequence—the primer binding site, or PBS—localized at the 3′ end of the U5 region. It has been suggested that the structure of the tRNAlys anti-codon loop is an important factor in its recognition by HIV-1 reverse transcriptase (46).

DNA synthesis proceeds up to the 5′ end of the RNA genome. Reverse transcriptase has an intrinsic RNaseH activity—the ability to degrade RNA in the context of a DNA/RNA hybrid. RNA degradation directly follows DNA polymerization, leaving 14 to 18 bases of complementary RNA and DNA double stranded. The resulting single-stranded DNA may jump to the 3′ terminus of the RNA molecule and complete minus strand DNA synthesis. Initiation of plus strand DNA synthesis is mediated by a polypurine RNA primer left intact because RNaseH does not cleave between purine residues. For most retroviruses, this polypurine tract (PPT) is located just to the 5′ end of U3. The double-stranded linear provirus is completed as shown in Fig. 4.

HIV, however, undergoes its genetic metamorphosis from RNA to DNA and back to RNA, with one nuance peculiar to lentiviruses and spumaviruses. Plus strand DNA synthesis is directed principally by two polypurine tracts (PPT) rather than one. The second PPT is located in the middle of the genome. DNA initiation is even more efficient from this site as opposed to the classical site just 3′ to the U3 region. In the unintegrated genome this replication strategy results in a single-stranded region in the center of the provirus. In fact, such a gapped structure was first described in 1981 for the visna virus (47,48). This partly single-stranded, partly double-stranded genome is reminiscent of the hepadna viruses. The final product is a double-stranded linear DNA that contains two copies of the LTR. The provirus is translocated to the nucleus where integration into the host genomic DNA occurs essentially randomly.

Organization of the Provirus

As Fig. 4 shows, replication results in two genetic metamorphoses: the reverse transcription of RNA into a double-stranded provirus and the reorganization of the genetic message. Thus, the U3 and U5 sequences are duplicated and reorganized. In

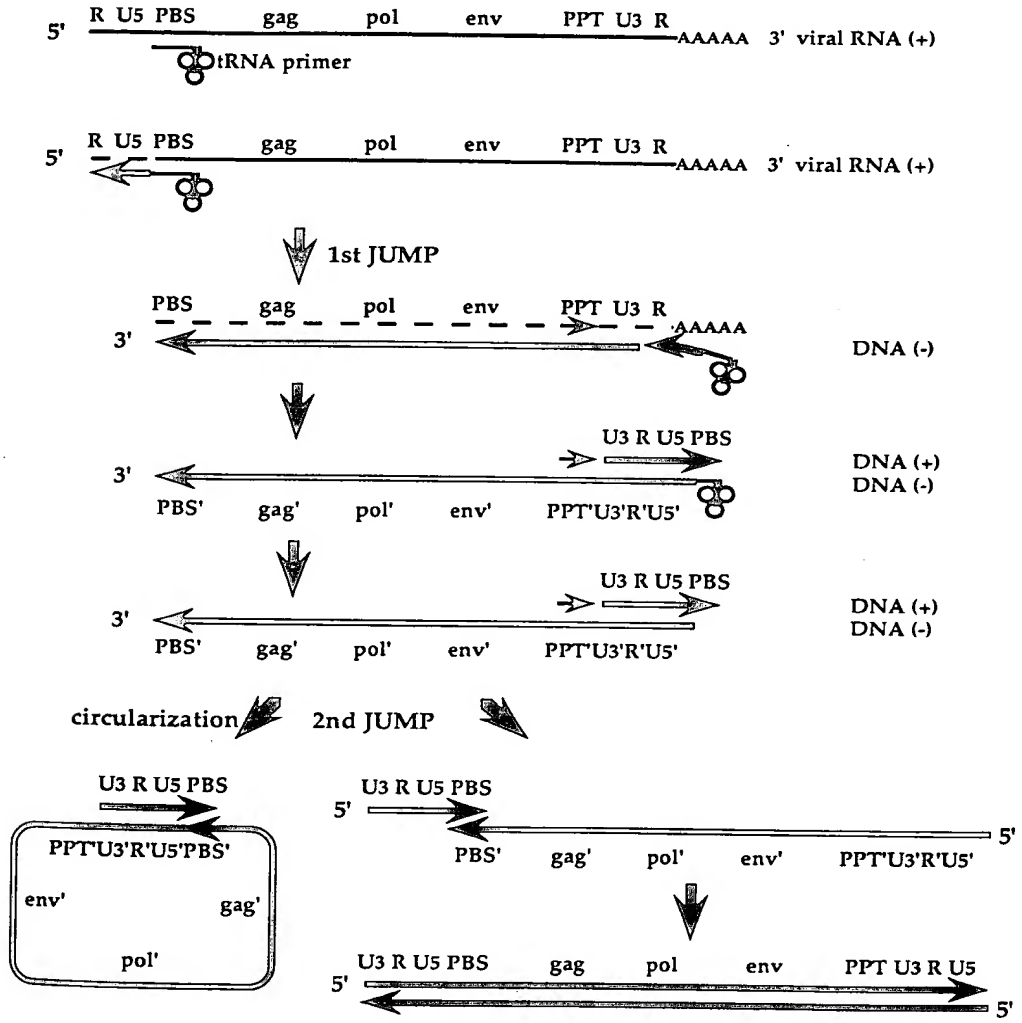

FIG. 4. Replication strategy of a retrovirus from the encapsidated RNA genome through to the double-stranded unintegrated provirus.

addition, only the 5′ copy of the R element of the LTR is reverse transcribed, although it is present twice in the RNA genome.

As with all retroviruses, the LTR structures of HIV harbor all the sequences necessary for proviral transcription and termination. However, the lentiviruses are complex in that, once transcribed, the fate of the full-length transcript is highly controlled, resulting in a temporal expression of certain proteins. (This extra layer of control, which involves sequences that map within coding sequences, will be dealt with in a subsequent section.) It is impossible to use "gene" in the normal sense when referring to retroviruses, because a single full-length transcript of the provirus is spliced into more than 20 mature mRNAs (49). In addition, at least two mRNAs

are bicistronic (50). For clarity, reference will be made to introns, exons, and open reading frames (orfs). An orf, which is part of one of three reading frames by which a ribosome can translate RNA, signifies an absence of stop codons and hence a protein coding potential. Between the two LTRs are nine orfs that encode 14 mature proteins.

There are four types of proteins (Table 3, Fig. 5): the viral structural proteins encoded by the *gag* and *env* orfs; the viral enzymes (*pol* orf); two proteins intimately associated with viral expression and its regulation, Tat and Rev; and a number of accessory proteins, notably Vif, Vpr, Vpu, and Nef.

Several secondary RNA structures are essential in the control of proviral expression, translation, and packaging. Furthermore, a large number of splice donor and acceptor sites overlap coding sequences. The importance of these features will be explored in subsequent sections.

REGULATION OF PROVIRAL TRANSCRIPTION

The transcription of the HIV provirus and the temporal ordering of splicing represent two profoundly original aspects of HIV virology. With hindsight it becomes clear that HIV has brilliantly solved two biologic conundrums. The first is how to survive in a lymphocyte that spends most of its time as a nondividing, resting lym-

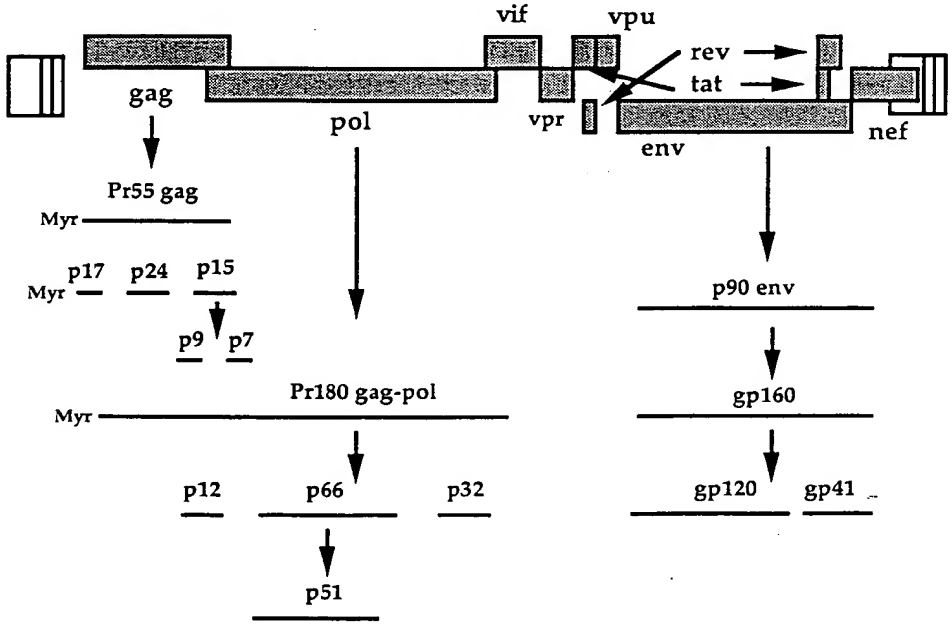

FIG. 5. HIV-1 proteins derived from polyprotein precursors. The Gag and Gag-Pol polyprotein precursors are cleaved by the viral proteinase p9. Cleavages of the envelope signal peptide and between gp120 and gp41 are the result of cellular enzymes. Myr indicates the myristylation at the amino terminus of the Gag and Gag-Pol polyproteins.

phocyte and only a small fraction of its time as an activated lymphoblast. The second is how to ensure the expression of nine orfs from one full-length genomic RNA. The solutions the virus uses for these events involve novel viral proteins called Tat and Rev.

Transcription Activation

For convenience, proviral transcription will be considered as starting from a latently infected lymphocyte. In a resting cell, it is presumed that there are no transcripts—presumed because it is virtually impossible to examine truly resting cells. Upon activation of a resting lymphocyte, the cell starts to produce many growth factors. The HIV LTR is susceptible to a number of these.

Transcription is initiated within the most 5′ LTR and terminates within the 3′ LTR. Transcription from the 3′ LTR is blocked by some unknown phenomenon. The 3′ LTR becomes functional, however, as soon as it is uncoupled from the 5′ LTR. The HIV-1 LTR is divided into three domains: U3, R, and U5. By definition, transcription initiation starts at nucleotide +1. All the *cis*-acting elements involved in HIV expression are localized in the U3 and R region. DNaseI footprinting of the HIV-1 LTR revealed four regions within the U3 R region that served as binding sites for numerous cellular proteins such as NF-κB, EBP-I, Sp1, CTF-1/NF-1, *TATA* binding protein, LBP-1, and UBP-1 (51,52). The target sequences for these proteins have been confirmed by mutagenesis and deletion studies. There are two copies of an NF-κB target sequence, or enhancer, and three of the target for the Sp1 factor.

The NF-κB enhancer sequence is perhaps a misnomer in that the two copies may be deleted from a functional provirus with little consequence. Deletion of the sequences from an isolated LTR would substantially reduce transcription, indicating the need for functional studies to be carried out on the complete provirus. Such studies indicated that the juxtaposed 3′ NF-κB and 5′ Sp1 sites were somehow intimately responsible for most of proviral transcription mediated by these two factors. Interestingly, the targets for CTF-1/NF-1 and LBP-1 beyond the transcription initiation site suggest that they may modulate the transcription complex more directly.

Other binding sites for eucaryotic transcription factors have been identified, including those for AP-1, NFAT-1, USF, NF-1, and SP50 (for reviews, see refs. 51 and 52). The sequence between −420 and −154 has been shown to contain a functional negative response element (NRE). The role of NRE in HIV replication has not been clearly defined.

The LTR may be activated *in vitro* by various T-cell mitogens such as phorbol esters, butyrate, interleukin 2, PHA, and PMA. Presumably the lymphocytes are activated principally by antigen *in vivo*. The LTR also may be activated by regulatory proteins of several DNA viruses including herpes simplex virus types 1 and 2, adenovirus, cytomegalovirus, hepadnavirus, pseudorabies virus, Epstein-Barr virus, human herpesvirus 6, and papovavirus. In addition, the transcription transactivator proteins Tax from HTLV-I and the spumaviral Bel 1 protein also may activate the HIV-1 LTR *in vitro*. Whether these phenomena are meaningful *in vivo* remains to be established. For these phenomena to be true, the silently infected resting T lymphocyte must be superinfected by a second virus, which is theoretically possible for HTLV, cytomegalovirus, and perhaps certain adenoviruses. Finally, ultraviolet irradiation also can induce HIV-1 transcription (53).

Once activated, the proviral transcription proceeds. Transcription is not efficient even in activated cells, with most transcripts terminating around +59 from the transcription start site, giving rise to a double-stranded RNA stem and loop structure called the *tat* response element (TAR). A small proportion goes beyond this point, resulting in full-length transcripts. These are rapidly spliced into small molecular weight mRNAs encoding principally *tat* and *rev,* two small yet complicated proteins whose function is still not completely understood.

Amplification of Transcription

Tat is a 86–100-residue protein of approximately 15.5 kDa; however, its first 56 residues are sufficient for full activity (54). Tat is a nuclear/nucleolar protein (55) that binds to a specific motif, called the bulge, in the TAR region of nascent RNA. A highly conserved cluster of seven cysteine residues has been implicated in dimerization via two zinc metal ions (56). The cysteine motif in the Tat protein differs significantly from the classic description of a zinc finger binding domain but is not unlike the cysteine-rich clusters of other metalloproteins. It is unclear whether this property is important for Tat function *in vivo,* as other reports have Tat active as a monomer (57).

Tat becomes a potent transactivator of viral gene expression via binding to the TAR region, a stem and loop structure in the R region of RNA. Both the location and orientation of TAR are critical for function (for reviews, see refs. 51 and 52). Whereas Tat binds to the bulge, a cellular protein is believed to bind to the loop. A number of phenomena then occur: the viral transcription initiation increases, the short transcripts elongate, and the initiation complexes stabilize. The increase in viral mRNA and protein synthesis stimulated by tat is greater than 100-fold in human HeLa cells or lymphocytes.

It has been suggested that transactivation by Tat is partly post-transcriptional. Using Xenopus oocytes, Tat was shown to activate the expression of presynthesized TAR mRNA despite the presence of transcriptional inhibitors (58). Yet the presence of TAR as a double-stranded RNA may inhibit the protein synthesis in *trans*. TAR can mediate the autophosphorylation and activation of the double-stranded RNA-dependent kinase (p68 kinase), which specifically can phosphorylate the eucaryotic initiation factor 2 (eiF-2), abolishing the mRNA translation. This area of research needs more elucidation.

The dominant role of Tat is as a transcriptional activator. Of the two proteins made from small (1.8–2 kb) mRNAs, Tat and Rev, Tat acts first. Within two to three hours of infection the level of Rev becomes sufficient for it to alter viral expression profoundly (59).

Coordination of Transcription

Rev is a small molecule of some 116 residues (19kDa) (60), of which the first 88 are sufficient for activity. Like Tat, Rev is a nuclear protein primarily found in the nucleolus. Rev binds tightly [Kd ~13nM] (61,62) to a complex RNA secondary/tertiary structure called the *rev*-responsive element (RRE), which is 234 bases long. As Fig. 5 shows, RRE is located within the *env* orf, or in the intron located at the 3′ end of the genome. A single Rev molecule may bind to RRE, resulting in multiple

activities. A number of domains include a highly basic sequence responsible for nuclear localization (63) and RNA binding. Another region of *rev* is involved in multimerization (64). Although Rev is phosphorylated at two serine residues, phosphorylation is not essential to *rev* function.

Because RRE is located in the most 3′ intron, rev cannot bind to the small mRNAs to influence their expression. By contrast, once bound to the larger RRE+ mRNAs, they can exit the nucleus into the cytoplasm, where they are promptly translated. The intermediate RRE+ mRNAs (4–5.5 kb) give rise to *vif, vpr, tat,* and *vpu+env* expression, and the largest 9-kb species—unspliced full-length genomic RNA—ensures expression of *gag* and *pol*. Many spliced donor and acceptor sequences are located within the central region of the genome. By alternative splicing, more than 20 distinct mRNAs have been described (49,54). Many differ by small noncoding leader exons. The major mRNAs are shown in Fig. 6.

Finally, an additional transcript that contains *tat, env,* and *rev* sequences and encodes a 28-kDa protein—Tev—has been described. Little is known about its role, however (65).

Note that the structural proteins Gag and Env and the Pol-derived enzymes are produced later in the cycle. Such temporal regulation of expression may be likened to the early and late phases of polyoma virus, adenovirus, or herpesvirus expression. Since this virus has a single operon, the Rev protein represents a means of achieving a balance between the expression of early and late transcripts.

Obviously, once Rev starts acting it has a positive effect on *gag, pol, vif, vpr, vpu,* and *env* expression as they are all made from RRE+ mRNAs (Fig. 6). Rev has a negative effect on its own expression and on that of *nef*. Large transcripts are made at the expense of small ones. Rev has little effect on *tat* expression, however, since there is a second *tat* mRNA, a 4.5-kb RRE+ species (Fig. 6).

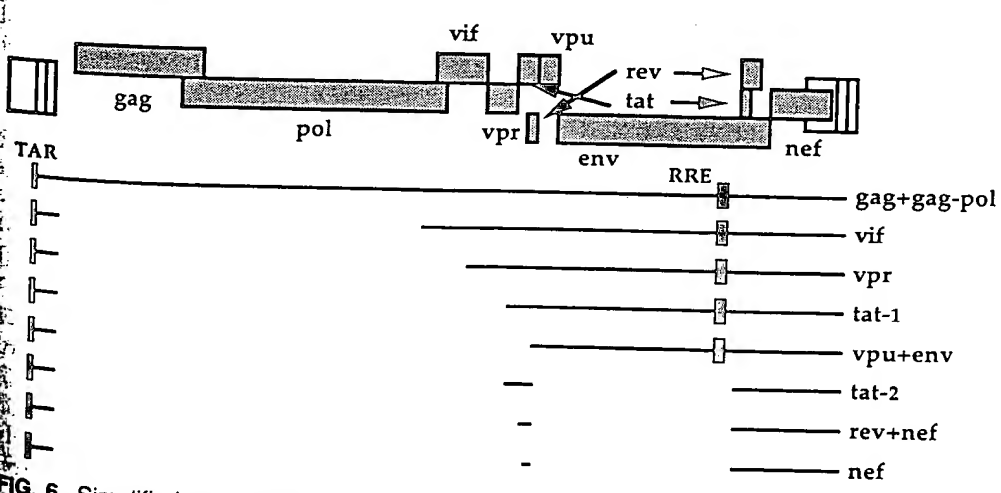

FIG. 6. Simplified transcription map for HIV-1. Only the major species have been shown. A number of small noncoding exons have been omitted for clarity. They map within the large 5′ intron as explained in the text. The Rev-dependent mRNAs are those bearing RRE. Adapted from ref. 49.

To summarize: first small transcripts are produced ensuring *tat*, *rev*, and *nef* expression. Two to three hours later there is enough Rev to ensure expression of the larger mRNAs (59). However, as *tat* is not down-regulated by Rev, the cell continues to churn out huge numbers of HIV transcripts, their fate being determined by the steady-state concentration of rev in the nucleus.

It is a source of amazement that a remarkably complex virus should derive so many different mRNAs from a single initial transcript and still survive. Perhaps the extraordinarily powerful Tat-driven transactivation of the genome results in so much viral RNA that even dividing it by nine (for each orf) results in sufficient amounts of each specific RNA.

The precise mechanism by which *rev* functions is not clear. It was suggested that Rev binding to RRE blocks splicing in the spliceosome (66). However, it is probably involved in transport between the nucleus and cytoplasm via the nucleolus. A recent report showed that human nucleolar B23 shuttle protein specifically binds *rev*. This association can be dissociated by RRE RNA *in vitro* and presumably in the nucleolus *in vivo* (67). It is as though *rev* has adapted to use a host cell nucleus/cytoplasm shuttle system. As Rev binds RRE in the nucleolus, B23 becomes displaced, and Rev/RRE+ mRNAs proceed into the cytoplasm. Rev then returns to the nucleus.

Bicistronic mRNAs

Two species of viral mRNA are bicistronic: the *vpu/env* and *rev/nef* mRNAs (50). In both cases the first initiator methionine (*vpu* or *rev*) is in a "weak" context for efficient translation whereas the second initiator methionine (*env* or *nef*) is in a "strong" context.

VIRION POLYPROTEINS

Gag Nucleocapsid Protein

The Gag polyprotein precursor is expressed from translation of full-length viral RNA. Pr55 gag is synthesized in the form of a 55-kDa precursor of approximately 500 amino acid residues. The NH_2 terminus being rapidly myristylated directs the precursor to the inner cytoplasmic membrane. Cleavage occurs during maturation of the complete immature virion and results in four proteins (Fig. 5). Pr55 Gag is cleaved uniquely by the HIV-1 aspartic protease. Both proteolytic cleavage and myristylation are required for the production of infectious virus particles. The NH_2 terminal protein, known as the p18 MA or matrix protein, forms the inner protein shell supporting the lipid bilayer. The major core protein, p24 capsid antigen (CA), self-aggregates to form the icosahedral inner shell within which the two RNA molecules are located. There are approximately 1,500 copies of CA per virion (68). The third Gag protein—p17 NA, or the nucleic acid binding protein—presumably is responsible for the condensation of two RNA genomes within the confined space of the virion core. P17 harbors a repetition of a zinc finger-like motif ($CX_2CX_4HX_4C$) and is highly basic. Although its mobility on an SDS gel is approximately 17 kDa, its real molecular weight is 9.5 kDa. It also may be cleaved in two, giving rise to p7 and p6 proteins. All three proteins—p17, p7, and p6—may be found in the virion. P7 retains the basic region and the cluster of cysteine residues. Mutations of the cysteine and

histidine residues in p17/p7 produce virions that are defective for packaging. The function of proline-rich p6 remains unclear, although mutations within block virus infectivity. Finally, defective HIV gag mutants of p17 can dominantly interfere with the replication of the wild-type virus (69), suggesting a novel possibility for gene therapy.

Pol-associated Enzymes

Like gag, pol is expressed from full-length genomic RNA. The principal viral enzymes are made as part of a 180-kDa gag-pol polyprotein precursor (Fig. 5). Pol is always expressed in concert with gag. This occurs within the region of overlap between *gag* and *pol* orfs by a −1 ribosomal frameshift during translation of genomic RNA. The site of the −1 frameshift is specific to a run of U residues and relies on an adjacent stem and loop structure—sometimes called a pseudoknot structure—for positioning (70,71). Its efficiency is approximately 10% to 15%; that is, 0.1 to 0.15 Gag-Pol polyproteins are synthesized for every Gag precursor. As the number of CA proteins has been estimated at approximately 1,500 per virion (68), the number of *pol*-derived molecules/virion is approximately 150 to 200. *Pol* encodes three distinct proteins: the protease, the reverse transcriptase/RNaseH, and the endonuclease/integrase.

The p9 protease, a member of the aspartic protease family, functions as a dimer. Within the immature virion monomeric protease, in the form of a Gag-Pol polyprotein precursor, dimerases result first in autocleavage and subsequent cleavage of other substrates. Cleavage of the Pr55 Gag and Pr190 Gag-pol proteins results in the morphologic changes termed maturation as seen by electron microscopy. The three-dimensional structure of the p9 protease was the first to be determined for the HIV proteins (72). It holds considerable promise for the design of anti-HIV reagents. As mentioned earlier, the HIV protease is an aspartic protease (73) and has a conserved pair of aspartic acid residues in the active site. Mutation of either results in immature and totally defective virions (74). Synthetic peptide analogues are potent inhibitors of purified HIV-1 protease.

The reverse transcriptase and RNaseH functions map to the NH_2 and COOH domains, respectively, of the p66 protein. Reverse transcriptase can exist as a p66 dimer or a p66:p51 heterodimer in which the RNaseH domain of one of the p66 subunits is cleaved away by the viral protease (75). Although both are functional in *in vitro* polymerization assays, the p66:p51 heterodimer is the most abundant form in the virion. Extensive mutagenesis studies have identified many key residues. Recently a high resolution 3.5Å crystal structure of the p66:p51 heterodimer has emerged (76). The DNA polymerization and RNaseH active sites are clearly distinguished (77,78), being separated by approximately 15 bases of duplex DNA, a value that supports the observations that RNaseH digestion of an RNA template succeeds polymerization by some 14 to 18 bases (79). Considerable conformational changes accompany duplex binding (78).

The p66 molecule comprises five distinct domains known as the finger, palm, thumb, connecting, and RNaseH domains. The p51 molecule is composed of only the first four. The most surprising finding is the remarkable structural asymmetry of the two subunits. Even though p51 is derived from p66, the conformation of the thumb and connecting domains is radically different. They effectively block the

DNA catalytic site of the p51 molecule. Thus, there is just one DNA catalytic site made up of a cleft formed by the finger, palm, and thumb domains of the p66 subunit. This loss of catalytic potential is more than offset by an increase in affinity of the heterodimer for the primer-template (80).

Those residues linked with resistance to RTase-inhibiting drugs map, not surprisingly, around the DNA catalytic and substrate binding sites. While the 3.5Å resolution was insufficient to place the side chain residues precisely, it is certain that such work will lead to the conception of more and more RTase inhibitors.

The most carboxyl-terminal of Pol proteins is the endonuclease/integrase, p32. This enzyme trims the ends of proviral DNA by two bases. The substrate for this is a covalently closed circular provirus with two LTRs. The form that is subsequently integrated, however, is a linear provirus devoid of the terminal dinucleotides. Chromosomal DNA is cleaved in a staggered fashion by the integrase, preferentially in the major groove on the surface of a nucleosome (81). Filling in results in a five-base pair duplication of cellular DNA.

Envelope Proteins

As the literature pertaining to the HIV-1 envelope is voluminous, only the most salient points will be developed here (for a review, see ref. 82). The envelope polyprotein precursor, gPr160 (Fig. 5), is heavily glycosylated. The protein precursor is progressively glycosylated as it passes through the endoplasmic reticulum, Golgi, and *trans*-Golgi, resulting in the addition of up to 70 kDa in carbohydrate. The amino terminal leader sequence is cleaved by a cellular protease associated with the docking protein complex. Cleavage into the mature surface and transmembrane proteins—gp120 and gp41—occurs by the time the protein has reached the *trans*-Golgi, probably by furin. Gp160, like gp120, is capable of binding the $CD4^+$ receptor, which can occur in the rough endoplasmic reticulum. The situation is apparently saved by the HIV-1 Vpu protein, which helps to transport gp160 out of the endoplasmic reticulum (83).

Gp41 remains anchored in the membrane. The long intracytoplasmic tail of 150 to 160 residues of gp41 is somewhat reminiscent of a cell surface receptor. It is present in most lentiviruses, with the exception of EIAV and FIV. Although the hydrophobic amino terminus of gp41 is involved in fusion between the virion and cellular membranes, the precise details are unknown. The first 129 amino acids of gp41 have been implicated in the assembly of Env protein tetramers (84). Indeed, cross-linking studies have suggested that the Env naturally exists on the virion surface as a tetramer.

The surface glycoprotein gp120, which is responsible for $CD4^+$ receptor recognition, is noncovalently associated with gp41, although shedding of the gp120 may occur within five hours of synthesis (85). Gp120 is a 470–490-residue protein harboring 22 cysteine residues, all of which form intramolecular disulfide bonds (Fig. 7) (82). Five hypervariable regions (V1-V5) are interspersed by constant regions. Four of the V regions represent disulfide bridged loops. Length varies considerably, particularly within the V1, V2, and V4 regions. There are 21 to 27 *N*-linked glycosylation sites, depending on the HIV-1 strain. For the single strain studied (86), all but one of the gp120 sites were occupied. Thus, gp120 is approximately 50% carbohydrate. The diverse structures of the side chains, those terminating in sialic acid, are unusual: 11 were mannose rich whereas the remaining 13 were complex glycans (86).

There are no *O*-linked sugars. Much has been written about the biologically important regions of gp120, although in the absence of a three-dimensional structure, caution must be exercised when interpreting these data. A number of biologically important regions have been mapped, including some involved in $CD4^+$ receptor or gp41 binding (for more details, see ref. 82).

One region of particular importance is the Gly-Pro-Gly (GPG) motif in the V3 loop (Fig. 7). Although it lies at the tip of a loop, the three residues are highly conserved. Studies have suggested that proteolytic cleavage just behind this motif is a prerequisite for cell-virion fusion. Research to date suggests that this motif is essential for the *in vitro* tropism of HIV. Thus mutation to GSG confers a B-cell tropism on the virus (87). If true *in vivo*, then it is particularly bad news, suggesting that HIV can broaden its tropism. A few naturally occurring HIV variants encoding this mutation

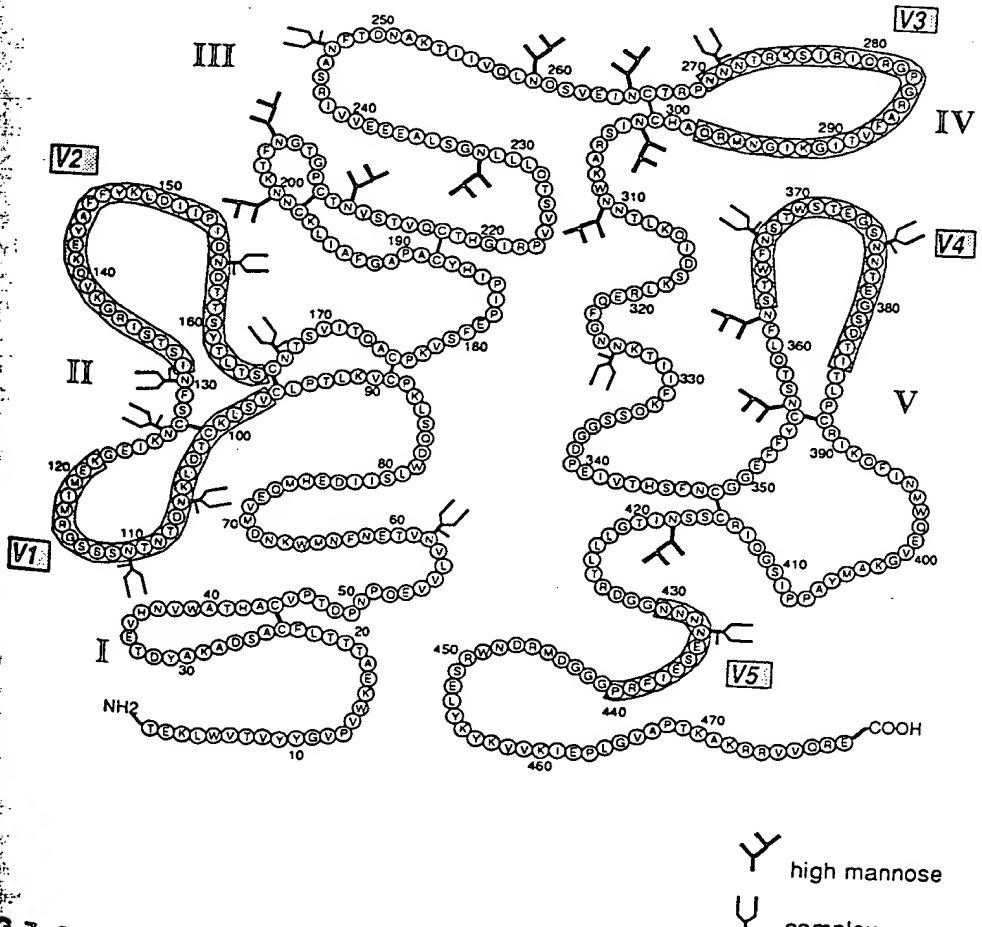

FIG. 7. Secondary structure and glycosylation pattern of HIV-1 gp120. The five hypervariable regions, V1-V5, are indicated. From ref. 86.

in vivo are now known. Other studies suggest that this region is important in defining the preference of a macrophage for lymphocyte-adapted virus (88,89). With all this work, the problem of interpreting the validity of using molecular-cloned virus derived from a tissue culture-adapted strain is again evident. The use of established transformed immature lymphocytic and monocytic cell lines further complicates the evaluation. The tremendous sequence variation inherent to gp120 will be addressed in a subsequent section.

ACCESSORY VIRAL PROTEINS

None of the accessory proteins for HIV-1—Nef, Vif, Vpr, or Vpu—is essential for replication. However *vif*-negative, *vpr*-negative, and *vpu*-negative viruses replicate to distinctly lower titers *in vitro*. *Nef*-negative mutants, although indistinguishable from wild-type virus *in vitro*, have profound effects *in vivo*. This helps explain the frequent finding of lesions in these orfs *in vitro*. In a few cases lesions were noted in all four accessory gene orfs (90).

The Vif protein is a 23-kDa protein (Table 3, Fig. 5) translated from a *rev*-dependent mRNA (Fig. 6). Although *vif* is an acronym for "viral infectivity factor," the mechanism of its action is unclear. *Vif*-negative mutants yield low levels of infectious viral particles (91). *Vif* does not seem to regulate viral gene expression or be involved in assembly of virions. *Vif* may interact with the cytoplasmic domain of gp41, perhaps via a cysteine protease activity (92).

Vpr is a conserved feature of all primate lentiviruses. The initial analysis of the SIV_{agm} sequence described the genome as *vpx*+ *vpr*− (93). It has been cogently argued, however, that based on a careful phylogenetic analysis of the sequences, this provirus actually encodes a *vpx*− *vpr*+ structure (94). In fact, it would appear that *vpx* is a highly divergent yet duplicated *vpr* orf. Although *vpr* is not essential for HIV-1 infectivity, replication, or cell killing, its exact function is not completely understood. Recently it has been suggested that *vpr* is an integral part of the virion (95), an intriguing hypothesis given that *vpx* is virion associated (96).

Vpu is a nonglycosylated protein of 80 to 82 residues (97) unique to the HIV-1 and SIV_{cpz} lentiviruses. Although these viruses are clearly related, there is just 36% sequence conservation between them (22). The amino terminus of p16 Vpu is hydrophobic and probably membrane associated. Vpu is also phosphorylated. *Vpu*-negative mutants decrease the amount of virus production in the culture supernatant. Virion morphology is highly irregular: diameter varies greatly and virions occasionally are binucleated. Despite this irregularity, the proportion of virus protein produced by such mutants is comparable to that of wild-type virus. It has been suggested that Vpu may function as a matrix protein at the level of virus assembly (98). Vpu is not incorporated into the virion, however. Recently it has been shown to promote dissociation of intracellular gp120/$CD4^+$ complexes (83).

A voluminous, but until recently inconclusive, literature accompanies p27 Nef. It is localized at the inner cytoplasmic membrane because of myristylation of the terminal glycine residue after elimination of the initiator methionine. GTP binding and GTPase activities proved fleeting, as did analogies with oncogenes (99). Although it was reported that Nef could be autophosphorylated or be a substrate for protein kinase C at threonine 15, many *nef* isolates lack this residue, suggesting that phosphorylation at this site may be gratuitous.

The acronym *nef*—for negative regulatory factor—derived from various studies that showed that *nef*-negative mutant viruses apparently replicated two- to ten-fold better than wild-type viruses. *Nef* was shown to down-regulate virus production through transcriptional inhibition of the LTR via the NRE. A number of well controlled studies have since been unable to substantiate these findings (59,100).

Two recent studies have helped to elucidate the role of *nef*. Using a retroviral vector it was shown that *nef* could induce down-regulation of cell-surface $CD4^+$ expression at a post-transcriptional level and by a pathway that is independent of $CD4^+$ serine phosphorylation (101). An *in vivo* study of SIV_{mac} *nef*-negative viruses derived from an infectious and pathogenic molecular clone showed that an irreversible deletion of 182 bases resulted in a much reduced viral load in experimentally infected animals. These animals did not develop AIDS, whereas the control animals did (102). Since $CD4^+$ is an accessory molecule to efficient HLA antigen presentation, the possibility that down-regulation of $CD4^+$ might help render the infected cell less vulnerable to cell-mediated killing may be considered.

Vpx, a 112-residue protein encoding an extraordinary run of seven proline residues at its COOH terminus, is unique to the HIV-2 and $SIV_{sm/mac}$ group of lentiviruses. Once again, *vpx*-negative mutants are replication competent although with reduced growth kinetics. The proportion of Vpx molecules to gag proteins encapsidated within the virion is approximately equimolar (96), indicating that it is an essential virion protein. Yet, HIV-1 and SIVs in African green monkeys (SIV_{agm}) and mandrills (SIV_{mnd}) replicate efficiently without *vpx*.

VIRION ASSEMBLY

HIV-1, like most retroviruses, is completely assembled at the inner cytoplasmic membrane. As noted above, the Pr55 Gag and Pr180 Gag-Pol precursors are myristylated at their amino terminus and are thus directed to the membrane. These precursors self-assemble (Fig. 7).

The viral envelope proteins also are expressed at the cell surface, but on the outer side. Gp120 remains noncovalently associated with the gp41 transmembrane protein. Thus, during the budding process, directed by gradual self-assembly of the Gag precursors, both host lipid and viral glycoprotein are incorporated. Gag and Env proteins may interact perhaps via the long cytoplasmic tail of gp41, but this has not been proved. Assembly is achieved without proteolytic cleavage of either of the Gag or Gag-Pol precursors. (For a review of the intricacies of viral assembly, see ref. 103.)

The arrangement of Gag proteins—the major core protein, the nucleic acid binding protein, and the matrix protein—within the Gag precursor ensures that they are where they should be within the virion, with the nucleic acid binding protein at the center, the major core protein surrounding the core, and the matrix protein surrounding the major core protein.

The packaging of two copies of full-length genomic RNA is achieved first by dimerization of two copies. The packaging signals are not all completely mapped. A secondary structure within the 5′ noncoding sequence has been implicated, as have sequences in the gag region for dimerization. The important feature is that these elements are found only in full-length RNA, whereas the smaller viral mRNAs, such as *vif* or *env*, lack these sequences. The nucleic acid binding protein may facilitate dimerization and packaging.

Once budded, the virion breaks away from the cell surface, leaving a complete yet immature virion. For reasons not yet known, the protease within the gag-pol precursor is activated, resulting in cleavage of the Gag and Gag-Pol precursors present in the immature virion. This cleavage results in the condensation of the core with the resulting characteristic morphology of the mature infectious virion. As mentioned earlier, maturation of the Gag precursors is essential for the production of infectious virions. The cycle is now complete.

HIV QUASISPECIES: GENETIC AND ANTIGENIC DIVERSITY

Retrovirus replication, indeed all RNA viral replication, proceeds in the absence of any proofreading mechanism. Consequently, polymerase errors, such as nucleotide misincorporations, go uncorrected. The average nucleotide misincorporation rates are of the order of 10^{-4} to 10^{-5}/base/replication cycle and are approximately one-million-fold greater than those associated with mammalian nuclear genes. RNA viral genomes are in general approximately $0.5–2.5 \times 10^4$ bases in length. A virus stock comprises a myriad of closely related viral genomes, most of which are genetically distinct (104). Thus, any genetic description of an RNA virus stock or isolate must take into account the inherently complex nature of the population of viral genomes that compose the whole.

These populations of viral genomes were termed "viral quasispecies" by Eigen et al., who described in mathematical terms some of the properties expected of such populations (105). The quasispecies are so complex and the RNA viruses in question so subtle that we can neither adequately describe these entities nor understand them (104). This is nowhere more evident than for HIV (26,106–114).

The challenge this diversity poses for biologic control of virus replication is daunting. From what we know of the copy numbers in an infected individual (30), each individual may have between 10^7 and 10^{11} proviruses, most of which are genetically distinct. The consequences for therapy and prevention are all too clear.

Before seroconversion, the HIV quasispecies tend to be homogeneous (26). Within two weeks of infection, however, considerable variation appears—certainly greater than the genetic variation seen during an acute phase influenza A infection (115) or even for flu sequences within an individual suffering from severe combined immunodeficiency (116). From work on the variation of SIV sequences with time, genetic complexity steadily increases after infection with virus of a single clone (117,118). In this respect the difference between influenza A and HIV is that the flu provokes an acute infection that is quickly resolved, as the individual has no natural reservoirs for the virus. As mentioned, acute-phase HIV infection is no different from acute-phase flu at the genetic level. However, as the HIV genome integrates into host cell DNA, it may sequester itself at a site of poor immunologic surveillance. Along with hiding itself away in resting lymphocytes, variant genomes may be generated and accumulated. Even though clearance of HIV virions may be efficient because of anti-HIV immune responses, elimination is not complete enough to prevent the accumulation of variants.

The genetic variation is so great that one study of *nef* sequences derived from PBMCs showed that there are as many *nef* variant sequences as clones sequenced. From these data it was estimated that any two proviruses *in vivo* (at least for the

samples in question) differed by at least 10 to 20 base substitutions (108). Even within a single organ, such as the spleen, the evidence of spatial inhomogeneity of HIV sequences was clear, not only in proviral copy number but also in sequence complexity (110). Furthermore, HIV variation is dynamic: HIV quasispecies may change in peripheral blood in as little as three months (106,107,111).

HIV quasispecies are sensitive to selection pressures (119). *In vitro* culture of HIV from PBMCs invariably results in the isolation of a subset of sequences present in the original population (107). Even continued passage of an isolate or passage to different established $CD4^+$ cell lines results in tremendous selection (120). The emergence of drug-resistant forms within as little as six months of treatment is another remarkable example of the effects of selection pressure (121,122).

Such extraordinary genetic diversity may not be necessary to AIDS pathogenesis, although some have suggested this to be the case (123). This diversity probably will not hamper our understanding of HIV gene function and replication, although caution should always be exercised in the interpretation of biologic significance. It becomes crucial as soon as one tries to intervene through virus culture, therapy, or prevention.

Numerous studies have shown that neutralizing antibody is of low titer and essentially isolate specific (82). Group-reactive neutralizing antibodies have been described but they appear to be of even lower titer. Neutralizing antibody is directed primarily to the V3 loop (Fig. 7) of the gp120 molecule. Although the tip of the loop is highly conserved, the adjacent sequences are highly variable (124). The V3 loop is a conformational epitope and neutralizing antibody escape mutants may result from changes within and without the loop. Frequently a single substitution is sufficient.

Long-term prospects for efficient combination chemotherapy may be brighter. If the frequency of producing a drug-resistant variant is of the order of 10^{-4}, then the probability of producing a variant against four or five independently acting drugs will be of the order of 10^{-16} to 10^{-20}. Yet the population of HIV proviruses in an infected individual is probably fewer than 10^{11}. Thus, by combining drugs, the probability of producing a resistant variant becomes so low as to be virtually impossible.

CONCLUSION

HIV, a redoubtable pathogen that will be with us for a long time, is a remarkably complex member of the lentiviral subfamily of retroviruses. *In vivo*, it infects T4 lymphocytes and antigen-presenting cells, either through the $CD4^+$ molecule or by Fc receptor–mediated antibody enhancement. Its replication cycle is efficiently adapted to that of the T4 lymphocyte: it remains transcriptionally silent in small resting cells. As soon as the T cell becomes activated, the powerful transactivation of the proviral genome leads to massive viral replication.

Retroviral replication proceeds in the absence of proofreading mechanisms. Consequently, replication is accompanied by the production of variant genomes. In the case of a single HIV-infected individual, the sheer size of variants is on the order of 10^7 to 10^{11}. This ensures an endless source of antibody escape and drug-resistant variants. Prevention and therapy strategies will have to confront this problem.

ACKNOWLEDGMENTS

We thank all our past and present collaborators. The literature is so vast that inevitably we could select only a few references from the many that merit citation; our apologies to our colleagues. This work was supported by grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA. JPV was supported by le Ministère de Recherche et de la Technologie.

REFERENCES

1. Martineau D, Bowser PR, Renshaw RR, Casey JW. Molecular characterization of a unique retrovirus associated with a fish tumor. *J Virol* 1992;66:596–599.
2. Fine D, Schochetman G. Type D primate retroviruses: a review. *Cancer Res* 1978;38:3123–3139.
3. Daniel MD, King NW, Letvin NL, Hunt RD, Sehgal PK, Desrosiers RC. A new type D retrovirus isolated from macaques with an immunodeficiency syndrome. *Science* 1984;223:602–605.
4. Weiss RA. Foamy retroviruses: a virus in search of a disease. *Nature* 1988;333:497–498.
5. Sonigo P, Barker C, Hunter E, Wain-Hobson S. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 1986;45:375–385.
6. Narayan O, Clements JE. Lentiviruses. In: Fields BN, Knipe DM, eds. *Virology*. New York: Raven Press, 1990;1571–1589.
7. Sigurdsson B, Thormar, Palsson PA. Cultivation of visna virus in tissue culture. *Arch Gesamte Virusforch* 1960;110:255–270.
8. Crawford TB, Adams DS, Cheevers WP, Cork LC. Chronic arthritis in goats caused by a retrovirus. *Science* 1980;207:997–999.
9. Kobayashi K. Studies on the cultivation of equine infectious anemia virus in vitro. II. Propagation of the virus in horse bone marrow culture. *Virus* 1961;11:189–201.
10. Van der Maaten MJ, Boothe AD, Seger CL. Isolation of a virus from cattle with persistent lymphocytosis. *J Natl Cancer Inst* 1972;49:1649–1657.
11. Barré-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983;220:868–871.
12. Daniel MD, Letvin NL, King NW, et al. Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* 1985;228:1201–1204.
13. Ohta Y, Masuda T, Tsujimoto H, et al. Isolation of simian immunodeficiency virus from African green monkeys and seroepidemiologic survey of the virus in various non-human primates. *Int J Cancer* 1988;41:115–122.
14. Tsujimoto H, Hasegawa A, Maki N, et al. Sequence of a novel simian immunodeficiency virus from a wild-caught African mandrill. *Nature* 1989;341:539–541.
15. Fultz PN, McClure HM, Anderson DC, Swenson RB, Anand R, Srinivasan A. Isolation of a T-lymphotropic retrovirus from naturally infected sooty mangabey monkeys (Cercocebus atys). *Proc Natl Acad Sci USA* 1986;83:5286–5290.
16. Murphey-Corb M, Martin LN, Rangan SR, et al. Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. *Nature* 1986;321:435–437.
17. Peeters M, Honore C, Huet T, et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* 1989;3:625–630.
18. Clavel F, Guétard D, Brun-Vézinet F, et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986;223:343–346.
19. Barin F, M'Boup S, Denis F, et al. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of West Africa. *Lancet* 1985;ii:1387–1389.
20. Pedersen NC, Ho EW, Brown ML, Yamamoto JK. Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science* 1987;235:790–793.
21. Haase AT. Pathogenesis of lentiviral infections. *Nature* 1986;322:130–136.
22. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* 1990;345:356–359.
23. Chakrabarti L, Guyader M, Alizon M, et al. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature* 1987;328:543–547.
24. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIV_{sm}) closely related to HIV-2. *Nature* 1989;339:389–392.
25. Ho DD, Moudgil T, Alam M. Quantitiation of human immunodeficiency virus type 1 in the blood of infected persons. *N Engl J Med* 1989;321:1621–1625.

26. Pang S, Shlesinger Y, Daar ES, Moudgil T, Ho DD, Chen IS. Rapid generation of sequence variation during primary HIV-1 infection. *AIDS* 1992;6:453–460.
27. Harper ME, Marselle LM, Gallo RC, Wong-Staal F. Detection of lymphocytes expressing human T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by in situ hybridization. *Proc Natl Acad Sci USA* 1986;83:772–776.
28. Schnittman SM, Psallidopoulos MC, Lane HC, et al. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. *Science* 1989;245:305–308.
29. Simmonds P. Variation in HIV virus load of individuals at different stages in infection: possible relationship with risk of transmission. *AIDS* 1990;4:S77-S83.
30. Schnittman SM, Lane HC, Greenhouse J, Justement JS, Baseler M, Fauci AS. Preferential infection of CD4+ memory T-cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. *Proc Natl Acad Sci USA* 1990;87:6058–6062
31. Harouse JM, Bhat S, Spitalnik SL, et al. Inhibition of entry of HIV-1 in neural cell lines by antibodies against galactosyl ceramide. *Science* 1991;253:320–323.
32. Dalgleish AG, Beverly PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984; 312:763–767.
33. Klatzmann D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 1984; 312:767–768.
34. Maddon PJ, Littman DR, Godfrey M, Maddon DE, Chess L, Axel R. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. *Cell* 1985; 42:93–104.
35. Lasky LA, Nakamura G, Smith DH, et al. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 1987;50:975–985.
36. Ryu SE, Kwong PD, Truneh A, et al. Crystal structure of an HIV-binding recombinant fragment of CD4. *Nature* 1990;348:419–426.
37. Wang JH, Yan YW, Garrett TP, et al. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 1990;348:411–418.
38. Peterson A, Seed B. Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. *Cell* 1988;54:65–72.
39. Arthos J, Deen CK, Chaikin MA, et al. Identification of the residues in human CD4 critical for the binding of HIV. *Cell* 1989;57:469–481.
40. Stein B, Gowda S, Lifson J, Penhallow R, Bensch K, Engelman E. pH independent HIV entry into CD4-positive T cells by human immunodeficiency virus. *Cell* 1987;49:659–668.
41. Pauza CD, Price TM. Human immunodeficiency virus infection of T cells and monocytes proceeds via receptor-mediated endocytosis. *J Cell Biol* 1988;107:959–968.
42. Zack JA, Arrigo SJ, Weitsman SR, Go AS, Haislip A, Chen IS. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. *Cell* 1990;61:213–222.
43. Homsy J, Meyer M, Tateno M, Clarkson S, Levy LA. The Fc and not CD4 receptor mediates antibody enhancement of HIV infection in human cells. *Science* 1989;244:1357–1360.
44. Takeda A, Tuazon CU, Ennis FA. Antibody-enhanced infection by HIV-1 via Fc receptor-mediated entry. *Science* 1988;242:580–583.
45. Bachelerie F, Alcami J, Arenzana-Seisdedos F, Virelizier JL. HIV enhancer activity perpetuated by NF-κB induction on infection of monocytes. *Nature* 1991;350:709–712.
46. Barat C, Lullien V, Schatz O, et al. HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA. *EMBO J* 1989;8:3279–3285.
47. Charneau P, Clavel F. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J Virol* 1991;65:2415–2421.
48. Harris JD, Scott JV, Traynor B, et al. Visna virus DNA: discovery of a novel gapped structure. *Virology* 1981;113:573–583.
49. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* 1990;64:2519–2529.
50. Schwartz S, Felber BK, Fenyo EM, Pavlakis GN. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J Virol* 1990;64:5448–5456.
51. Cullen BR. Human immunodeficiency virus as a prototypic complex retrovirus. *J Virol* 1991;65:1053–1056.
52. Cann AJ, Karn J. Molecular biology of HIV: new insights into the virus life-cycle. *AIDS* 1989;3(suppl 1):S19-S34.
53. Valerie K, Delers A, Bruck C, et al. Activation of human immunodeficiency virus type 1 by DNA damage in human cells. *Nature* 1988;333:78–81.

54. Muesing MA, Smith DH, Capon DJ. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell* 1987;48:691–701.
55. Ruben S, Perkins A, Purcell R, et al. Structural and functional characterization of human immunodeficiency virus tat protein. *J Virol* 1989;63:1–8.
56. Frankel AD, Bredt DS, Pabo CO. Tat protein from human immunodeficiency virus forms a metal-linked dimer. *Science* 1988;240:70–73.
57. Dingwall C, Ernberg I, Gait MJ, et al. Human immunodeficiency virus 1 tat protein binds trans-activation responsive region (TAR) RNA in vitro. *Proc Natl Acad Sci USA* 1989;86:6925–6929.
58. Braddock M, Thorburn AM, Chambers A, et al. A nuclear translational block imposed by the HIV-1 U3 region is relieved by the Tat-TAR interaction. *Cell* 1990;62:1123–1133.
59. Kim SY, Byrn R, Groopman J, Baltimore D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol* 1989;63:3708–3713.
60. Sodroski J, Goh WC, Rosen C, Dayton A, Terwilliger E, Haseltine W. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature* 1986;321:412–417.
61. Heaphy S, Dingwall C, Ernberg I, et al. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* 1990;60:685–693.
62. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans- activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 1989;338:254–257.
63. Cochrane AW, Golub E, Volsky D, Ruben S, Rosen CA. Functional significance of phosphorylation to the human immunodeficiency virus Rev protein. *J Virol* 1989;63:4438–4440.
64. Malim MH, McCarn DF, Tiley LS, Cullen BR. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J Virol* 1991;65:4248–4254.
65. Benko DM, Schwartz S, Pavlakis GN, Felber BK. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env and rev proteins. *J Virol* 1990;64:2505–2518.
66. Chang DD, Sharp PA. Regulation by HIV Rev depends upon recognition of splice sites. *Cell* 1989;59:789–795.
67. Fankhauser C, Izaurralde E, Adachi Y, Wingfield P, Laemmli UK. Specific complex of human immunodeficiency virus type 1 rev and nucleolar B23 proteins: dissociation by the Rev response element. *Mol Cell Biol* 1991;11:2567–2575.
68. Luftig R, Ikuta K, Bu M, Calkins P. Terminal stages of retrovirus morphogenesis. In:Pearl LE, ed. *Retroviral proteinases control of maturation and morphogenesis*. London: Macmillan Press, 1990;141–148.
69. Trono D, Feinberg MB, Baltimore D. HIV-1 Gag mutants can dominantly interfere with the replication of the wild-type virus. *Cell* 1989;59:113–120.
70. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 1988;331:280–283.
71. Jacks T. Translational suppression in gene expression in retroviruses and retrotransposons. *Curr Top Microbiol Immunol* 1990;157:93–124.
72. Navia MA, Fitzgerald PMD, McKeever BM, et al. Three-dimensional structure of aspartyl-protease from human immonodeficiency virus HIV-1. *Nature* 1989;337:615–620.
73. Oroszlan S, Luftig RB. Retroviral proteinases. *Curr Top Microbiol Immunol* 1990;157:153–185.
74. Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci USA* 1989;86:5781–5785.
75. Ferris AL, Hizi A, Showalter SD, et al. Immunologic and proteolytic analysis of HIV-1 reverse transcriptase structure. *Virology* 1990;175:456–464.
76. Kohlstaedt LA, Wang J, Friedman JM, Jordan SR, Matthews DA. Crystal structure at 3.5Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 1992;256:1783–1790.
77. Davies JF 2d, Hostomska Z, Hostomsky Z, Jordan SR, Matthews DA. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* 1991;252:88–95.
78. Arnold E, Jacobo-Molina A, Nanni RG, et al. Structure of HIV-1 reverse transcriptase/DNA complex at 7Å resolution showing active site locations. *Nature* 1992;357:85–89.
79. Fu TB, Taylor J. When retroviral reverse transcriptases reach the end of their templates. *J Virol* 1992;66:4271–4278.
80. Anderson SF, Coleman JE. Conformational changes of HIV reverse transcriptase subunits on formation of the heterodimer: correlation with kcat and Km. *Biochem* 1992;31:8221–8228.
81. Pryciak PM, Varmus HE. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 1992;69:769–780.
82. Weiss RA. Cellular receptors and viral glycoproteins involved in retrovirus entry. In: Levy JA, ed. *The Retroviruses*, vol 2. New York: Plenum Press, (*in press*).

83. Willey RL, Maldarelli F, Martin MA, Strebel K. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J Virol* 1992;66:226–234.
84. Earl PL, Doms RW, Moss B. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc Natl Acad Sci USA* 1990;87:648–652.
85. Kieny MP, Rautmann G, Schmitt D, et al. AIDS virus env protein expressed from a recombinant vaccinia virus. *Biotechnology* 1986;4:790–795.
86. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* 1990;265:10373–10382.
87. Takeuchi Y, Akutsu M, Murayama K, Shimizu N, Hoshino H. Host-range mutant of human immunodeficiency type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. *J Virol* 1991;65:1710–1718.
88. Cheng-Mayer C, Seto D, Tateno M, Levy A. Biologic features of HIV-1 that correlate with virulence in the host. *Science* 1988;240:80–82.
89. O'Brien WA, Koyanagi Y, Namazie A, et al. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* 1990;348:69–73.
90. Ratner L, Haseltine W, Patarca R, et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 1985;313:277–284.
91. Luciw PA, Cheng-Mayer C, Levy JA. Mutational analysis of the human immunodeficiency virus: the orf-B region down-regulates virus replication. *Proc Natl Acad Sci USA* 1987;84:1434–1438.
92. Guy B, Geist M, Dott K, Spehner D, Kieny MP, Lecocq J-P. A specific inhibitor of cysteine proteases impairs a Vif-dependent modification of human immunodeficiency virus type 1 Env protein. *J Virol* 1991;65:1325–1331.
93. Fukasawa M, Miura T, Hasegawa A, et al. Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV-SIV group. *Nature* 1988;333:457–461.
94. Tristem M, Marshall C, Karpas A, Petrik J, Hill F. Origin of vpx in lentiviruses. *Nature* 1990;347:341–342.
95. Cohen EA, Dehni G, Sodroski JG, Haseltine WA. Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J Virol* 1990;64:3097–3099.
96. Henderson LE, Sowde RC, Copeland TD, Benveniste RE, Oroszlan S. Isolation and characterization of a novel protein (X-ORF product) from SIV and HIV-2. *Science* 1988;241:199–201.
97. Cohen EA, Terwilliger EF, Sodroski JG, Haseltine WA. Identification of a protein encoded by the vpu gene of HIV-1. *Nature* 1988;334:532–534.
98. Strebel K, Klimkait T, Maldarelli F, Martin MA. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J Virol* 1989;63:3784–3791.
99. Guy B, Kieny MP, Riviere Y, et al. HIV F/3′ orf encodes a phosphorylated GTP-binding protein resembling an oncogene product. *Nature* 1987;330:266–269.
100. Hammes SR, Dixon EP, Malim MH, Cullen BR, Greene WC. Nef protein of human immunodeficiency virus type 1: evidence against its role as a transcriptional inhibitor. *Proc Natl Acad Sci USA* 1989;86:9549–9553.
101. Garcia JV, Miller AD. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* 1991;350:508–511.
102. Kestler HW III, Ringler DJ, Mori K, et al. Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* 1991;65:651–662.
103. Gelderblom HR. Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS* 1991;5:617–637.
104. Holland JJ, De La Torre JC, Steinhauer DA. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 1992;176:1–20.
105. Eigen M, Biebricher CK. Sequence space and quasispecies distribution. In: Domingo E, Holland JJ, Ahlquist P, eds. *RNA Genetics*, Boca Raton: CRC Press Inc. 1988;211–245.
106. Wain-Hobson S. HIV quasispecies in vivo and in vitro. *Curr Top Microbiol Immunol* 1992;176:181–193.
107. Meyerhans A, Cheynier R, Albert J, et al. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* 1989;58:901–910.
108. Delassus S, Cheynier R, Wain-Hobson S. Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro. *J Virol* 1991;65:225–231.
109. Delassus S, Meyerhans A, Cheynier R, Wain-Hobson S. Absence of selection of HIV-1 variants in vivo based on transcription/transactivation during progression to AIDS. *Virology* 1992;188:811–818.
110. Delassus S, Cheynier R, Wain-Hobson S. Nonhomogenous distribution of human immunodeficiency virus type 1 proviruses in the spleen. *J Virol* 1992;66:5642–5645.

111. Martins LP, Chenciner N, Asjo B, Meyerhans A, Wain-Hobson S. Independent fluctuation of human immunodeficiency virus type 1 rev and gp41 quasispecies in vivo. *J Virol* 1991;65:4502–4507.
112. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990;64:5840–5850.
113. Epstein LG, Kuiken C, Blumberg BM, et al. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* 1991;180:583–590.
114. Pang S, Vinters HV, Akashi T, O'Brien WA, Chen ISY. HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. *AIDS* 1991;4:1082–1092.
115. Robertson JS, Nicolson C, Bootman JS, Major D, Robertson EW, Wood JM. Sequence analysis of the haemagglutinin (HA) of influenza A (H1N1) viruses present in clinical material and comparison with the HA of laboratory-derived virus. *J Gen Virol* 1991;72: 2671–2677.
116. Rocha E, Cox NJ, Black RA, Harmon MW, Harrison CJ, Kendal AP. Antigenic and genetic variation in influenza A (H1N1) virus isolates recovered from a persistently infected immunodeficient child. *J Virol* 1991;65:2340–2350.
117. Burns DP, Desrosiers RC. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. *J Virol* 1991;65:1843–1854.
118. Johnson PR, Hamm TE, Goldstein S, Kitov S, Hirsch VM. The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* 1991;185:217–228.
119. Domingo E. RNA virus evolution and the control of viral disease. *Prog Drug Res* 1989;33:93–133.
120. Vartanian JP, Meyerhans A, Åsjö B, Wain-Hobson S. Selection, recombination and G—>A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol* 1991;65:1779–1788.
121. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 1988;246:1155–1158.
122. St Clair MH, Martin JL, Tudor-Williams G, et al. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science* 1992;253:1557–1559.
123. Nowak MA, Anderson RM, McLean AR, Wolfs TF, Goudsmit J, May RM. Antigenic diversity thresholds and the development of AIDS. *Science* 1991;254:963–969.
124. Myers G, Rabson AB, Berzofsky JA, Smith TF, Wong-Staal F. *Human Retroviruses and AIDS*. Los Alamos: Los Alamos National Laboratory, 1991.
125. Doolittle RF, Feng DF, McClure MA, Johnson MS. Retrovirus phylogeny and evolution. *Curr Top Microbiol Immunol* 1990;157:1–18.